The silver gene of *Drosophila melanogaster* encodes multiple carboxypeptidases similar to mammalian prohormone-processing enzymes

(polyprotein/neuropeptide/melanization/svr gene)

STEPHEN H. SETTLE, JR.*, M. M. GREEN, AND KENNETH C. BURTIS[†]

Section of Molecular and Cellular Biology, University of California, Davis, CA 95616

Contributed by M. M. Green, June 29, 1995

The silver (svr) gene of Drosophila melano-ABSTRACT gaster is required for viability, and severe mutant alleles result in death prior to eclosion. Adult flies homozygous or hemizygous for weaker alleles display several visible phenotypes, including cuticular structures that are pale and silvery in color due to reduced melanization. We have identified and cloned the DNA encoding the svr gene and determined the sequence of several partially overlapping cDNAs derived from svr mRNAs. The predicted amino acid sequence of the polypeptides encoded by these cDNAs indicates that the silver proteins are members of the family of preprotein-processing carboxypeptidases that includes the human carboxypeptidases E, M, and N. One class of svr mRNAs is alternatively spliced to encode at least two polyproteins, each of which is composed of two carboxypeptidase domains.

The silver (*svr*) gene, which maps near the distal end of the *Drosophila* X chromosome, is one of numerous genes required for the proper melanization and sclerotization of the insect cuticle (1). The phenotypes associated with mutations in the *svr* gene include effects not only on pigmentation but also on viability, wing differentiation, catecholamine catabolite pools, and behavioral responses to light (1). One approach to furthering our understanding of the complex pathways involved in these processes is to molecularly characterize the genes involved. Here we report on the cloning and characterization of the *svr* gene[‡], first identified by Mohr in 1918 and rediscovered by Bridges in 1921 (2).

MATERIALS AND METHODS

Isolation of Terminal Deficiencies. Females homozygous for the third-chromosome mutation mu-2 are defective in their ability to repair double-stranded chromosome breaks (3). mu-2females were x-irradiated (1000 roentgens) and crossed to males carrying a Y chromosome bearing a yellow² (y^2) marked duplication of the distal tip of the X chromosome. Male F₁ progeny receiving a terminally deleted maternal X chromosome were identified by their y^2 phenotype. Because the y gene is located distal to *svr*, such deletions were further tested genetically to determine whether or not they removed more proximal genes, including l(1)EC5, *svr*, and *elav*.

Molecular Biological Methods. Preparation of plasmid and phage DNA was by standard methods (4). Northern blot analysis and DNA sequencing were done as previously published (5). *P*-element transformation (6) used the vector pCaSpeR 4 (7). Sequences were analyzed with the University of Wisconsin Genetics Computer Group suite of programs (8) and BLAST algorithms on the World Wide Web server at the National Center for Biotechnology Information.

RESULTS

Localization of the *svr* **Gene.** Genetic mapping of the *svr* gene determined its location to the distal region of the X chromosome (2). The cytological location of *svr* (polytene band 1B7) is intriguing in that the gene lies amidst a cluster of genes involved in nervous system development, including the achaete-scute complex, the *elav* gene and the l(1)vnd gene (9). The previously cloned *elav* gene (10) provided a starting point for cloning the *svr* gene, which is the next identified gene distal to *elav*.

The approximate location of the *svr* gene was determined by deletion analysis (11). These studies determined that the order of genes (distal to proximal) in this region was scute, l(1)EC4, $\{l(1)EC5, svr\}$, *elav*, *vnd*. Because no terminal deficiency was available that defined a boundary separating l(1)EC5 and *svr*, we carried out a screen for new terminal deficiencies failing to complement l(1)EC5 but complementing *svr*, or vice versa.

Terminally deficient X chromosomes were generated as described in *Materials and Methods*. Two X chromosomes thus produced, $y^{-(1)}$ and $y^{-(2)}$, were found by genetic complementation tests to be $l(1)EC5^-$ and svr^+ , suggesting the presence of a terminal break between these two genes. Polytene cytology confirmed the existence of terminal deficiencies with breakpoints in approximately the correct locations (data not shown).

A unique feature of one of the two *mu-2*-derived terminal deficiencies enabled us to serendipitously determine the distal boundary of the *svr* gene with greater precision than that allowed by mapping the original breakpoint. The ends of such terminal deficiencies are unstable and display a continuous loss of terminal sequences, at a rate of \approx 75 bp per generation (12). During a 1-year period (20-25 generations) after the isolation of the $Df(1)y^{-(1)}$ chromosome, it lost its previous ability to complement *svr* mutations. We hypothesize that this loss was the result of the terminal deficiency beginning to delete sequences essential for *svr* function, thus defining the distal boundary of the *svr* gene to the \approx 2-kb region distal to the molecular position of the breakpoint determined in the *svr*⁻ chromosome.

The molecular location of the $Df(1)y^{-(1)}$ breakpoint was determined by Southern blot analysis (Figs. 1 and 2). Probes for this analysis were radiolabeled restriction fragments from λ phage clones isolated in a chromosomal walk in the *elav-vnd* region (10). The region examined included the 35 kb of DNA situated distal to the previously mapped location of the *elav* gene; probes distal to this region were not tested, as they failed to hybridize *in situ* to polytene chromosome squashes of the $Df(1)y^{-(1)}$ chromosome (data not shown). Southern blot anal-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

^{*}Present address: Department of Developmental Biology, Stanford Medical School, Stanford, CA 94305.

[†]To whom reprint requests should be addressed.

[‡]The sequences reported in this paper have been deposited in the GenBank data base (accession nos. U29591 and U29592).



FIG. 1. Mutations affecting the size of restriction fragments in the *svr* gene. Southern blot of *Eco*RI-digested DNA from either *svr*^{73L12} (lane 1), *svr*^{poi} (lane 2), wild-type (lane 3), or $Df(1)y^{-(1)}$ (lanes 4 and 5) flies was probed with the 6.5-, 2.7-, and 1.5-kb *Eco*RI fragments (Fig. 2). Size of hybridizing bands is indicated.

ysis of DNA from terminal deficiency-bearing flies revealed a decrease in the size of a single EcoRI fragment (6.5 kb in wild type; 4.9 kb in the mutant chromosome) from this region (Fig. 1, compare lane 3 with lanes 4 and 5). A reduction of 2.0 kb in the size of a 13.4-kb *Bam*HI fragment from the same region was also associated with the terminal deficiency (data not shown).

Two lines of evidence suggest that these size polymorphisms are the result of a terminal deletion event in the $Df(1)y^{-(1)}$ chromosome. Both polymorphisms are consistent with a deletion breakpoint removing 1.6 kb from the distal end of the 6.5-kb *Eco*RI fragment (this results in a loss of 2.0 kb from the 13.4-kb *Bam*HI fragment). Furthermore, the polymorphic fragments migrate as fuzzy bands (compare the 4.9-kb and 6.5-kb bands in Fig. 1, lane 4), consistent with the stochastic nature of DNA loss over many generations from terminally deleted chromosomes (12). Thus, it is likely that the distal boundary of sequences essential for *svr* function is located between coordinates +2 and +4 on the map of the *svr* region presented in Fig. 2.

Additional evidence correlating svr function with sequences in this region was obtained by Southern blot analysis of DNA from flies homozygous for two viable svr alleles, svr^{73L12} (isolated by M.M.G.) and svrpoi (13). Both mutations are associated with alterations in the electrophoretic mobility of EcoRI fragments located proximal to the fragment containing the $Df(1)y^{-(1)}$ breakpoint. The pattern seen in DNA from svr^{73L12} mutants is consistent with the insertion of a 7.3-kb transposable element (with an internal EcoRI site) into the 2.7-kb EcoRI fragment (Fig. 1, lane 1), whereas the pattern seen with svr^{poi} is consistent with a 0.9-kb insertion into the 1.5-kb EcoRI fragment lying immediately proximal to the 2.7-kb fragment (Fig. 1, lane 2; Fig. 2). Additional restriction fragment length polymorphisms are also seen in the DNA from this region in the different mutant strains (data not shown). However, in conjunction with data from Northern blot analysis (see below), these findings support the hypothesis that these aberrations are associated with svr gene.



FIG. 3. Northern blot of mRNAs from adults (lanes 1–3), 0- to 12-hr embryos (even-numbered lanes 4–16), and 12- to 24-hr embryos (odd-numbered lanes 5–17). Lanes 1–3 included RNA from wild-type, svr^1 , or svr^{poi} flies, respectively, and were hybridized with a mixed probe including the 6.5-, 2.7-, and 1.5-kb *Eco*RI fragments. Embryonic mRNA blots were hybridized with probes from different restriction fragments, including the 14.5-kb (lanes 4 and 5), 6.5-kb (lanes 6 and 7), 2.7-kb (lanes 8 and 9), 1.5-kb (lanes 10 and 11), 8.3-kb (lanes 12 and 13), 2.4-kb (lanes 14 and 15), and 4.8-kb (lanes 16 and 17) *Eco*RI fragments.

A complex set of mRNAs is transcribed from the region surrounding the svr mutation breakpoints. Northern blots of RNA from 0- to 12-hr and 12- to 24-hr embryos (Fig. 3, lanes 4-17) revealed mRNAs of 1.5 kb, 2.5 kb, and 3.3 kb, as well as a set of mRNAs 6-7 kb in length. Adjacent restriction fragments detected these different size classes of mRNAs as follows: 6.5-kb fragment (1.5, 2.5, 3.3, 6-7 kb); 2.7-kb fragment (1.5, 2.5, 3.3, 6-7 kb); 1.5-kb fragment (2.5 and 3.3 kb faintly, 6-7 kb); and 8.3-kb fragment (6-7 kb only). Northern blots probed with fragments of cDNA clones confirmed that all of these transcripts shared common sequences, and the use of strand-specific probes indicated that transcription was from distal to proximal (data not shown). A large EcoRI fragment (14.5 kb) distal of the 6.5-kb fragment also detected multiple transcripts, although it is unclear whether these are related to the family of transcripts described above. In mRNA from pupae the set of transcripts detected by this distal region appeared to comprise a slightly shorter version of the set of mRNAs found in embryos, consistent with the possibility of an alternative (smaller) first exon located in this region that is spliced onto common exons located downstream (data not shown). As previously reported (10), the 8.3-kb fragment as well as the more proximal 2.4- and 4.8-kb fragments each hybridized (Fig. 3, lanes 12-17) to mRNAs encoded by elav. Thus, the svr and elav transcription units are most likely adjacent to one another, although the structure of the small transcript detected by the 8.3-kb fragment (lane 13) is still unclear.

The correlation of *svr* gene expression with the family of transcripts described above is strengthened by the observation that production of these transcripts is altered in flies homozygous for the mutant allele *svr^{poi}*, which is associated with aberrations in the DNA sequence of the 1.5-kb *Eco*RI fragment. In *svr^{poi}* flies the large, 6- to 7-kb class of mRNAs is completely absent (Fig. 3, lane 3). Preliminary evidence also suggests that the *svr^{73L12}* mutation results in significantly reduced levels of *svr* mRNAs (data not shown), although it is



FIG. 2. Map of the *svr* region. Indicated (from top) are orientation relative to chromosome, polarity and approximate location of the *svr* and *elav* transcribed regions, location and size of *Eco*RI fragments, extent of $Df(1)y^{-(1)}$ deletion (heavy black arrow), location of *Bam*HI fragment sufficient for rescue (crosshatched bar), and scale in kilobases.

not certain whether the transcripts produced are of normal structure.

Final confirmation of the identity of sequences sufficient to provide *svr* function was obtained by showing that a 13.4-kb *Bam*HI fragment of genomic DNA (Fig. 2), including the entirety of the 6.5-, 2.7-, and 1.5-kb *Eco*RI fragments as well as flanking sequences from the next most proximal and distal *Eco*RI fragments, was able to completely rescue the lethal phenotype associated with several alleles of *svr*, when introduced into the *Drosophila* genome by *P* element-mediated germline transformation (data not shown). The visible pigmentation phenotype was only partially rescued by these constructs and may require downstream sequences not present in the 13.4-kb *Bam*HI fragment.

Sequence Analysis of the svr cDNAs. The 6.5-, 2.7-, and 1.5-kb fragments of genomic DNA from the svr region were used as probes to screen for cDNA clones in a λ ZAP (Stratagene) library prepared from *Drosophila* embryonic mRNA. Three phagemid clones (3, 6, and 15A) were rescued from hybridizing phage for further characterization.

The three cDNAs fell into two classes with respect to their patterns of hybridization to Southern blots of genomic DNA fragments from the *svr* region (data not shown). cDNA 6 hybridized to the 6.5-, 2.7-, and 1.5-kb EcoRI fragments (Fig. 2), as well as to a 3.5-kb sequence that includes the proximal half of the 6.5-kb fragment. It did not hybridize to the 9.0-kb EcoRI fragment. cDNA 3 and cDNA 15A each hybridized to the 6.5-, 2.5-, 1.5-, and 9.0-kb EcoRI fragments, but not to the 3.5-kb fragment mentioned above. Subsequent DNA sequence analysis revealed that cDNA 6 was truncated proximally at a native EcoRI site separating the 2.5-kb and 9.0-kb genomic fragments and thus was most likely derived from a cDNA that originally extended into the 9.0-kb fragment.

Because each of the three cDNAs contains (or is deduced to have originally contained, as discussed above) sequences from the 9.0-kb genomic fragment, and since Northern blot results indicate that sequences from the 9.0-kb fragment are included in only the large, 6- to 7-kb size class of mRNAs, we presume that the cDNAs studied are derived from one or more of the large mRNAs. The structure of the smaller transcripts is thus uncertain at this time. None of the three cDNAs represents a complete copy of the large transcript, but a composite sequence of 3956 nt can be derived by combining sequences from the three overlapping clones (GenBank accession no. U29592). Although not a complete copy of a *svr* mRNA, this composite sequence includes a single large open reading frame encoding a polypeptide of 1119 amino acids with a predicted M_r of 126,498.

Because the Southern blot results indicated a difference between cDNA 15A and cDNA 6 with respect to hybridization to the distal and proximal ends of the 6.5-kb genomic EcoRIfragment, the sequence of the first 207 nt of cDNA 15A was determined (GenBank accession no. U29591). This sequence spanned the location of the alternative splice junction differentiating these two cDNAs and revealed the functional consequences of this splice with respect to the coding sequence (see below). However, cDNA 15A is clearly truncated at its 5' end and does not extend to an initiation codon. Thus, the amino-terminal sequence of the polypeptide encoded by this class of mRNAs remains to be determined.

The svr Gene Encodes Tandem Carboxypeptidases Expressed as Polyproteins. When the polypeptide encoded by the composite cDNA was compared with a nonredundant set of databases by using the FASTP algorithm on the BLAST network server of the National Center for Biotechnology Information, two domains closely related to the vertebrate nondigestive carboxypeptidases M, N, and E (14–16) were revealed. These two domains each comprise a complete carboxypeptidase as judged from comparison with the vertebrate sequences. Interestingly, an intron separates the sequences encoding the first

and second domains (data not shown; ref. 17). The svr gene most likely encodes a third carboxypeptidase domain unique from the first two, based on the 58-amino acid sequence predicted from the sequence of the alternative 5' exon present in the truncated cDNA clone 15A. Because the complete sequence of this region is not yet available, it cannot be stated with certainty that there is a third complete domain. If there were, the alternative splicing event would result in an mRNA encoding a polyprotein in which the amino-terminal part of the first carboxypeptidase domain would be unique, with the remainder shared in common between the two splice forms. We have designated these three putative carboxypeptidases as CbpDm-Ib (the amino-terminal domain encoded by the composite cDNA), CbpDm-II (the carboxyl-terminal domain), and CbpDm-Ia (the putative carboxypeptidase based on the partial sequence from cDNA 15A). An alignment of these three polypeptide sequences with the human carboxypeptidases is shown in Fig. 4.

Remarkably, the carboxyl-terminal portion of the *svr*encoded polypeptide (residues 842–1119) comprises yet a fourth highly diverged carboxypeptidase domain (referred to as CbpDm-III in Fig. 4). Although this domain is truncated with respect to the full-length carboxypeptidase sequences by a termination codon, there is significant conservation of amino acid sequence in several regions. It is highly unlikely that this fourth domain has enzymatic activity, as discussed below.

The sequence of a *Drosophila* genomic DNA fragment encoding part of a carboxypeptidase E homologue was recently reported (ref. 17; GenBank accession no. U03883). This fragment is clearly identical to the 1.5-kb *Eco*RI fragment of the *svr* gene (Fig. 2) and encodes residues 425–695 of the predicted *svr* polypeptide, comprising slightly less than the amino-terminal half of the second (CbpDm-II) carboxypeptidase region.

It is not possible from sequence alignments to make an unambiguous correlation between the three Drosophila carboxypeptidase domains and carboxypeptidases N, E, and M. Evaluations of similarity depend significantly on which specific regions are chosen for analysis. However, it is clear that carboxypeptidase domains CbpDm-Ib and CbpDm-II have diverged significantly from each other, and that at many positions each is more similar to one of the vertebrate enzymes than it is to the other Drosophila sequence. By most criteria, CbpDm-Ia is more similar to CbpDm-Ib than to any of the other enzymes, but it is still significantly different, and a final analysis must await the determination of the remainder of the sequence. Ultimately, biochemical and immunolocalization experiments will be essential for determining which of these enzymes, if any, is the likeliest functional homolog for specific vertebrate carboxypeptidases.

Inspection of the predicted amino acid sequence reveals several relevant features in common between the Drosophila protein and the vertebrate carboxypeptidases, including a signal peptide with predicted cleavage site (18) after Gly²⁵. Furthermore, each of several highly conserved residues with roles in the enzymatic activity of carboxypeptidase (14) is present in the CbpDm-Ib and CbpDm-II sequences (Fig. 4). The conserved histidine at position 101 in CbpDm-1b is replaced by a glutamine in CbpDm-1a, although the adjacent conserved residues are present. CbpDm-Ia protein may represent a novel form of carboxypeptidase, it may be nonfunctional, or the single nucleotide change responsible for this alteration may be the result of an error during reverse transcription of cDNA 15A. Further experiments are needed to discern which of these possibilities is correct. None of the highly conserved residues is present in the very diverged CbpDm-III carboxyl-terminal domain, although both of the two regions of CbpDm-III with the greatest similarity to the other carboxypeptidases are located immediately adjacent to

			10	20	30
CbpDm-1b CpbDm-II	•••••	mptlgll	fasigiavla	mgvphcrgyt	ikedesflqq afnafltptk
CbpDm-III				ddpqhwasan	dfriienvVn
Cope Copm	magrggsall	alcgalaacg	lgLllplvaa	df	rriqqedgis
Copn		.msdllsvfl	h1L11		fklvapVt
CONS			D		v-
ChoDm-1b	40	50 difaciekav	60 PromisionflG	70	80 Laterntrer
CpbDm-II	YehHnFtaME	sYLraIsssY	PSLTRLYSIG	KSVqGRdLWV	LEIfatPGsH
CbpDm-III	trYHtnpqvr FeyhBypel.r	arLaeLengn	gqIas.Fgya	dSefGtifny	LkMTsdiGep
Clopm	.nYHRqEgME	aFLktVagnY	ssVThLhSIG	KSVkGRnLWV	LvVgrfPkeH
Coppn	FrhHRYDdLv	rtLykVqnec	PgITRVYSIG	RSVEGRhLyV	LEfSdhPGiH
CONS	I-IIIKIE-DB		F-11KV1516	KJVLKJK-LWV	1615FG-n
ChroDm-1b	90 nlltPovKYI	100 ANMHGdEt.V	110 GRaLLVYMAa	120 YLlonberis	130 D.LOOLVINST
CpbDm-II	VPGVPEFKYV	ANMHGNEV.V	GKELLLiLtk	YMlErYgn.d	DrItkLVNgT
CbpDm-III CbpDm-Ia	EehkykiivV	SSLYDITADL ANIGGDEA.V	GREILLILI	hLvEgFklqd YLathYdodo	tsVVeLLkrS k.VcaLLNlT
Cbpe	EPGePEFKYI	gNMHGNEA.V	GRELLIFLAG	YLcneYqkgn	EtIVnLIhST
Clopm Clopn	riGiPEFKYV EPlePEvKYV	ANMHGOLET.V	GRELLLALID	YLvtsdgkdp	E.ItnLINST arIVaLIadT
CONS	EPG-PEFKYV	ANMHGNEA-V	GRELLLYLAE	YL-E-Y	E-IV-LINST
	140	1,50	160	170	
CbpDm-1b	divivpTM	NPDGYalsge	gnceslpnyv	GRONAANIDL	NRdFPD
CboDm-III	vivFLPatSk	fonvFDmvns	ntsicdo.vL	GdelAerIl.	NRNFPDQ
CbpDm-Ia	eIHFLPTc	NPDGFakAk.			
Cope	RIHIMPSL	NPDGFERAAS	pdcvvsI	GRENVNGVDL	NRNFPDIGF1
Copn	RIHILP SM	NPDGYEvAaa	qgpnkpgylV	GRnNAnGVDL	NRNFPDInty
COINS	RIHFLPSM	NPDGFE-A	V	GR-NA-GIUL	*
Chorn-1b		180	190	200	210
CpbDm-II	.YgtDrf		nkvt	EPEvaAVMnW	tlslpfvlsa
CbpDm-III Choe		nnhl lkNm	gpet	DqakdvfLqf	LrSerFdLml
Cope	·····	afeyNn	vsr	qPETvAVMkW	LktetFVLSA
Copn	iYynEkyggp	nhhlplpdNw	ksqv	EPETrAVIrW	MhSfnFVLSA
CONS	-16			EFEI-AVI-W	1-3-77 1134
ChroDm-1b	220	230 VPVDnsl.ah	necc	240 eeS1T	250 PDDrVFkaLA
CobDm-II	NLHGGsLVAN	YPFDdnendf	ndpfmrlrns	singrkpnpT	eDnaLFkhLA
CbpDm-III Choe	tfgaGnsdlN NLHGGdLVAN	YP	· · · · · · · · · · · · · · · · · · ·	sasahevSsS	kgDsVlekfA PDDalFOsLA
Clopm	NLHGGALVAS	YPFDngvqat	galy	srSlT	PDDdVFQyLA
Coppn CONS	NLHGGAVVAN NLHGGALVAN	YPYDks	fehrvr	gvrrtasTpT S-T	PDDkLFQkLA PDD-VFO-LA
	* 260		270	200	200
CbpDm-1b	260 HtYSdnHPiM	rkGnnC	.nDsFsgG	280 ITNGAhWYeL	SGGMQDFNYa
CpbDm-II	giYSnaHPtM	ylGqpCelfq	neFFPDG	ITNGAGWYSV	TGGMQDwNYV
Copum-111 Cope	raySsfnPaM	sdpnrppcrk	ndDdssFvDG	tTNGgaWYSV	pGGMQDFNYL
Clopm	HtYasrnPnM	kkGdeC	. knkmnFPnG	VINGySWYpL	QGGMQDYNYI
CONS	H-YSHP-M	GC	dyFPDG	ITNGASWISL	SCGMQDFNYL
	300	310	320	* 330	340
CbpDm-1b	fsNCFEL	TIELSCCKYP	aastLpgeWg	rNKaSLLQ1L	rQaHiGIKGl
CpbDm-II CbpDm-IJT	ragClEL	TLaISCCRmP	kaaeLsrYWe hgkkIaSFga	anRepLLQFI ktsirLrtFw	EQVHCGIhGf 1
Cbpe	SSNCFEI	TVELSCeKFP	pEEtLkTYWe	dNKnSLIsYL	EQIHrGVKGf
Copm Copn	waqCFEI	TLELSCCKYP TLELSCCKYP	rEEkLpSFWn	nNKaSLIEYI oNReaLIOFL	kQVH1GVKGq FOVHaGIKGm
CONS	NCFEI	TLELSCCKFP	-EE-L-S-W-	-NK-SLIQFL	EQVH-GIKG-
	350	360	370	380	390
CbpDm-1b	VtDasGfPIA	dAnVyVaGLe	ek.pmrTsKr	GEYWRLLLEPG	lysVhASAfG
Cbpe	VrDlqGNPIA	NAtisVeGid	HdvtSqvf	GDYWRLLIPG	nYkLTASAPG
Clopm	VfDonGNPLD	NvIVeVqdrk	HicpyrTnKy	GEYYLLLPG	SYIINVTVPG
CONS	V-DGNPIA	NAVV-V-Gi-	HS-K-	GDYWRLLLPG	-Y-vtAsAPG
	400	410	418		
CbpDm-1b	Yqtsapqqvr	vtndnqealr	ldfkLaPvet	nfdgnfrkvk	verseppq
CpbDm-II Chne	YaP1rmeVeV Ylaitkkvev	Polyhpfemrm	dItlMp	eekeelmow	kmmsetlnf
Clopm	hdPhitkViI	Peksqnfsal	kkdiLlPfqg	qldsipvsnp	scpmiplyrn
Clopn CONS	YdPetvtVtV Y-PV-V	gpaeptlvnf P	hLkrsiPqvs -1]-P	pvrrapsrrh	gvrakvqpqa
Copn	rkkememral	grgpa	niir		

FIG. 4. Amino acid sequences of the polypeptide encoded by the composite *svr* cDNA sequence, separated into the two complete (CbpDm-Ib and CbpDm-II) and one partial (CbpDm-III) carboxypeptidase domains as well as the partial amino-terminal domain encoded by the alternative exon represented by cDNA 15A (CbpDm-Ia) are aligned with the sequences of the human carboxypeptidases E, M, and N (Protein Identification Resource accession nos. S09489, A32619, and S02074, respectively). The consensus sequence (CONS) represents positions at which there were at least three occurrences of similar (lowercase in consensus) or identical (uppercase in consensus) amino acid residues (similar residues include D/E, I/L/M/V, K/R, S/T, and

the positions defined by conserved residues Glu^{104} and Glu^{305} in the CbpDm-Ib domain.

None of the previously known nondigestive carboxypeptidases are expressed as polyproteins, although there is evidence for posttranslational cleavage of at least one, carboxypeptidase E (16). Thus, it would seem unlikely that the *Drosophila* proteins remain in their dimeric state after synthesis. There is a consensus cleavage site, Lys-Leu-Lys-Lys at positions 440– 443 (19, 20), for the subtilisin/kexin family of serine endoproteases located between CbpDm-Ib and CbpDm-II. Thus, Cbp-Dm-Ib (or CbpDm-Ia) may be cleaved posttranslationally from CbpDm-II and perhaps autocatalytically remove the terminal lysine residues from CbpDm-Ib (or CbpDm-Ia).

Polyprotein enzymes are extremely unusual in nonviral higher eukaryotic animal systems, although there are examples of structural polyproteins, such as ubiquitin and profilaggrin (21, 22). It may be significant from an evolutionary perspective that the one group of animal polypeptides commonly found as part of polyproteins, the neuropeptide hormones, are joined in this category by one of the enzymes that processes them.

Since a significant portion of CbpDm-II is encoded in the 9.0-kb fragment, it is likely that the smaller transcripts, which are presumably terminated or polyadenylylated in sequences arising from the 1.5-kb fragment, encode either CbpDm-Ia or CbpDm-Ib. Isolation and characterization of polyadenylylated cDNAs representing the smaller mRNAs will be necessary to precisely determine their coding potential.

DISCUSSION

Carboxypeptidases play an essential role in the processing of numerous neuropeptide hormones, as well as other bioactive peptides and proteins. These exopeptidases function by removing one or two arginine or lysine residues from the carboxyl-terminal ends remaining after endoproteolytic cleavage of proprotein precursors, resulting in activation, inactivation, or modulation of the biological activity of the substrate peptides. Different members of the family of proproteinprocessing carboxypeptidases are expressed in a tissue-specific manner in higher vertebrates, each serving a specific role: carboxypeptidase E functions in the processing of peptide hormones and neurotransmitters in secretory vesicles; carboxypeptidase M acts as a membrane bound protease on the surface of many cell types, potentially to modify the structure and activity of precursor peptides before or after they interact with cell surface receptors; and carboxypeptidase N is secreted into the plasma by the liver, where it serves to inactivate peptides such as anaphylatoxins and kinins released into the circulatory system.

The svr gene potentially encodes three alternative carboxypeptidases. Despite this intriguing numerical coincidence, no firm correlation can be drawn between these and the three vertebrate enzymes on the basis of amino acid sequence homologies, since sequence comparisons between the vertebrate enzymes and different regions of the svr-encoded protein lead to different conclusions. Nonetheless, it is clear that Drosophila can express more than one type of carboxypeptidase and that these enzymes differ in sequence from each other as significantly as they do from the vertebrate enzymes. Further investigation of the enzymatic properties and cellular

F/Y). Amino acid residues that are identical or similar to the consensus appear in uppercase type. Numbers indicate the position of amino acid residues in the CbpDm-1b domain, beginning at the amino terminus of the *svr* polypeptide. Highly conserved residues critical for carboxypeptidase activity are indicated by stars below the consensus sequence. Putative endopeptidase cleavage site between domains CbpDm-1b and CbpDm-II is underlined at beginning of CbpDm-II sequence. GenBank accession nos. for the silver sequences are U29591 and U29592.

distribution of the various forms will be required to determine whether the *Drosophila* enzymes share functional homologies with specific vertebrate counterparts.

There is a clear biological requirement for a carboxypeptidase E-type activity in *Drosophila*, given the large number of neuropeptides known to modulate various physiological processes in insects. The recent cloning from *Drosophila* or other insect species of the genes encoding many of these neuropeptides (23–34) makes it possible to identify the likely sites of endoproteolytic cleavage of the precursors by the wellcharacterized preprotein convertases of the subtilisin/kexin family (35). In all cases except for prothoracicotropic hormone and eclosion hormone, in which the mature neuropeptide comprises the carboxyl-terminal portion of the precursor, it is clear that a carboxypeptidase E-like activity is essential to remove the terminal basic residues from the prohormone.

In vertebrate systems, expression of carboxypeptidase E is strongly correlated with sites of neuropeptide synthesis (36). Preliminary Northern blot data (S.H.S., unpublished work) indicates that *Drosophila* heads are significantly enriched in *svr* mRNAs with respect to bodies, as would be expected for expression in the adult central nervous system. The location of the small number of specific neurons in the *Drosophila* central nervous system that interact with antibodies against several neuropeptides has been determined (37), and it will be of interest to examine the possibility of a correlation between the expression of the various carboxypeptidases encoded by the *svr* gene and neurons expressing specific prohormone precursors.

The phenotype of *svr^{poi}* mutants, which lack the 6- to 7-kb size class of mRNAs, suggests that the CbpDm-II carboxypeptidase, while playing a role in melanization and wing development, is not essential for viability. This may be indicative of a specialized role for each of the *Drosophila* carboxypeptidases in specific physiological processes or, alternatively, may reflect functional redundancy of the enzymes, with any of the forms providing the minimal function required for viability.

The large number of potential neuropeptide targets for the svr carboxypeptidases is consistent with the pleiotropic phenotype of some hypomorphic svr alleles and the lethal effect of other alleles. Strong alleles would be expected to be lethal due to catastrophic effects on neuropeptide maturation, particularly if the svr gene were the only source of this type of carboxypeptidase. Specific phenotypic effects on pigmentation, wing shape, catecholamine pools, and photonegativity all may reflect different alterations in neuropeptide processing, although it is not yet possible to specify any particular causal relationships. It is tempting to speculate that defects in processing a Drosophila homolog of neuropeptides controlling melanization, such as those found in other insects (38, 39), might be responsible for the reduced pigmentation seen in svr mutant flies. However, neither the identity nor the role of melanizing hormones in Drosophila has been established. The characterization of these and other potential neuropeptide targets of carboxypeptidase processing will be essential for fully understanding the role of the svr gene in Drosophila development.

We thank Kalpana White and colleagues for generously providing phage clones from the *elav* region, Brett Enlow for assistance with DNA sequencing, and Bruce Black for advice and encouragement along the way. This work was supported by National Science Foundation Grant DMB-8916620. S.H.S. was supported by a University of California President's Undergraduate Fellowship.

- 1. Wright, T. R. (1987) Adv. Genet. 24, 127-222.
- Lindsley, D. L. & Grell, E. H. (1968) Genetic Variations of Drosophila melanogaster (Carnegie Inst. Washington, Baltimore), Vol. 627.

- Mason, J. M., Strobel, E. & Green, M. M. (1984) Proc. Natl. Acad. Sci. USA 81, 6090-6094.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
- 5. Burtis, K. & Baker, B. (1989) Cell 56, 997-1010.
- 6. Rubin, G. M. & Spradling, A. C. (1982) Science 218, 348-353.
- Pirrotta, V. (1988) Vectors for P-element Transformation in Drosophila, eds. Rodriguez, R. L. & Denhardt, D. T. (Butterworths, Boston), pp. 437-456.
- Devereux, J., Haeberli, P. & Smithies, O. (1984) Nucleic Acids Res. 12, 387-395.
- Jimenez, F. & Campos-Ortega, J. (1987) J. Neurogenet. 4, 179– 200.
- Campos, A. R., Rosen, D. R., Robinow, S. N. & White, K. (1987) EMBO J. 6, 425–431.
- Mason, J. M., Voelker, R. A., Rosen, D., Campos, A. R., White, K. & Lim, J. K. (1986) Dros. Info. Serv. 63, 164.
- Biessmann, H., Carter, S. B. & Mason, J. M. (1990) Proc. Natl. Acad. Sci. USA 87, 1758-1761.
- 13. Goldschmidt, R. B. (1944) Drosoph. Inf. Serv. 18, 40-44.
- Tan, F., Chan, S. J., Steiner, D. F., Schilling, J. W. & Skidgel, R. A. (1989) J. Biol. Chem. 264, 13165-13170.
- Gebhard, W., Schube, M. & Eulitz, M. (1989) Eur. J. Biochem. 178, 603–607.
- Manser, E., Fernandez, D., Loo, L., Goh, P. Y., Monfries, C., Hall, C. & Lim, L. (1990) *Biochem. J.* 267, 517–525.
- 17. Bernasconi, P. (1994) Arch. Insect Biochem. Physiol. 27, 169-178.
- 18. von Heijne, G. (1986) Nucleic Acids Res. 14, 4683-4690.
- Hosaka, M., Nagahama, M., Kim, W. S., Watanabe, T., Hatsuzawa, K., Ikemizu, J., Murakami, K. & Nakayama, K. (1991) J. Biol. Chem. 266, 12127-12130.
- Stone, T. E., Li, J. P. & Bernasconi, P. (1994) Arch. Insect Biochem. Physiol. 27, 193-203.
- Wiborg, O., Pedersen, M. S., Wind, A., Berglund, L. E., Marcker, K. A. & Vuust, J. (1985) EMBO J. 4, 755-759.
- 22. Gan, S. Q., McBride, O. W., Idler, W. W., Markova, N. & Steinert, P. M. (1990) *Biochemistry* 29, 9432-9440.
- Bradfield, J. Y. & Keeley, L. L. (1989) J. Biol. Chem. 264, 12791-12793.
- Donly, B. C., Ding, Q., Tobe, S. S. & Bendena, W. G. (1993) Proc. Natl. Acad. Sci. USA 90, 8807–8811.
- Adachi, T., Takiya, S., Suzuki, Y., Iwami, M., Kawakami, A., Takahashi, S. Y., Ishizaki, H., Nagasawa, H. & Suzuki, A. (1989) *J. Biol. Chem.* 264, 7681-7685.
- Sato, Y., Oguchi, M., Menjo, N., Imai, K., Saito, H., Ikeda, M., Isobe, M. & Yamashita, O. (1993) Proc. Natl. Acad. Sci. USA 90, 3251-3255.
- Digan, M. E., Roberts, D. N., Enderlin, F. E., Woodworth, A. R. & Kramer, S. J. (1992) Proc. Natl. Acad. Sci. USA 89, 11074– 11078.
- Davis, M. T., Vakharia, V. N., Henry, J., Kempe, T. G. & Raina, A. K. (1992) Proc. Natl. Acad. Sci. USA 89, 142–146.
- Kawakami, A., Kataoka, H., Oka, T., Mizoguchi, A., Kimura-Kawakami, M., Adachi, T., Iwamin, M., Nagasawa, H., Suzuki, A. & Ishizaki, H. (1990) Science 247, 1333–1335.
- 30. Veenstra, J. A. (1994) Biochem. Biophys. Res. Commun. 204, 292-296.
- Nichols, R., Schneuwly, S. A. & Dixon, J. E. (1988) J. Biol. Chem. 263, 12167–12170.
- Horodyski, F. M., Riddiford, L. M. & Truman, J. W. (1989) Proc. Natl. Acad. Sci. USA 86, 8123–8127.
- Schneider, L. E. & Taghert, P. H. (1988) Proc. Natl. Acad. Sci. USA 85, 1993–1997.
- 34. Feany, M. B. & Quinn, W. G. (1995) Science 268, 869-873.
- Seidah, N. G. & Chretien, M. (1994) Methods Enzymol. 244, 175-188.
- 36. Fricker, L. D. (1988) Annu. Rev. Physiol. 50, 309-321.
- 37. Zitnan, D., Sehnal, F. & Bryant, P. J. (1993) Dev. Biol. 156, 117-135.
- Matsumoto, S., Isogai, A., Suzuki, A., Ogura, N. & Sonobe, H. (1981) Insect Biochem. 11, 725–733.
- Hiruma, K., Norman, A. & Riddiford, L. M. (1993) J. Insect Physiol. 39, 353-360.